Paper Recommendation System Based on Paper Citation Network
Group -11

Sezal Chug (2017101)   **Priya Kaushal (2017081)**   **Raghav Rathi (2017083)**

Git Repo Link: https://github.com/raghav17083/Collaborative-Filtering

AIM

The research presents a collaborative filtering based recommendation approach for scientific papers that does not depend on priori user profiles and which utilizes only public contextual information.

1. Using citation context, we utilized various matrix completion methods to find hidden associations between papers.
2. To evaluate the performance of the proposed approach we used publicly available dataset- ACL Anthology network
3. The experimental results are measured in terms of precision and recall, which are commonly used information retrieval metrics.

The novelty of this study is that, with the proposed approach, researchers are able to find relevant and useful publications over the internet regardless of their previous research experiences and research area.

Dataset Description

ACL Anthology Network (AAN): A manually curated networked database of citations, collaborations, and summaries in the field of Computational Linguistics.
Files Used
1. Acl-metadata.txt
    a. paperID, paperTitle, paperAuthor, paperVenue, paperYear
2. Paper_ids.txt
    a. paperID, paperTitle, paperYear
3. Paper-citation-network.txt
    a. paperIDs of citing and cited paper p1 ⇒ p2

Methodology

A description of all algorithms used in the project are as follows:

1. User Paper Based
2. Co-Cited and Co-Referenced Based Similarity
3. KMeans
4. Kernel K Means

5. Subspace Clustering
6. Latent Factor Model
7. Nuclear Norm Minimization
8. Binary matrix Factorisation
9. CCIDF- Co-citation Inverse Document Frequency

Results

Using the **Citation Re-Prediction Evaluation Methodology,** we calculate the recall and precision values for each POI to get recommendations and respective graphs

1. <u>User Paper Based</u>
    a. **Papers Recommended for Paper ID- P10-1142**
    Title- Supervised Noun Phrase Coreference Research: The First Fifteen Years
    1 . Supervised Models for Coreference Resolution   D09-1101
    2 . Graph-Cut-Based Anaphoricity Determination for Coreference Resolution   N09-1065
    3 . Unsupervised Models for Coreference Resolution   D08-1067
    4 . A Global Relaxation Labeling Approach to Coreference Resolution   C10-2125
    5 . Coreference Resolution with World Knowledge   P11-1082
    b. **Papers Recommended for Paper ID- P12-1041**
    Title- Coreference Semantics from Web Features
    1 . End-to-End Coreference Resolution via Hypergraph Partitioning   C10-1017
    2 . State-of-the-art NLP Approaches to Coreference Resolution: Theory and Practical Recipes   P09-5006
    3 . Coreference Resolution in a Modular Entity-Centered Model   N10-1061
    4 . Learning Structured Perceptrons for Coreference Resolution with Latent Antecedents and Non-local Features   P14-1005
    5 . Extracting Bacteria Biotopes with Semi-supervised Named Entity Recognition and Coreference Resolution   W11-1814

2. <u>Co-Cited and Co-Referenced Based Similarity</u>
    a. **Papers Recommended for Paper ID- P10-1142**
    Title- Supervised Noun Phrase Coreference Research: The First Fifteen Years-
    1 . Unsupervised Learning of Narrative Event Chains   P08-1090
    2 . An Empirically Based System For Processing Definite Descriptions   J00-4003

3 . CogNIAC: High Precision Coreference With Limited Knowledge And Linguistic Resources   W97-1306

4 . Error-Driven Analysis of Challenges in Coreference Resolution   D13-1027

5 . Coreference Resolution across Corpora: Languages Coding Schemes and Preprocessing Information   P10-1144

b. Papers Recommended for Paper ID- P12-1041

Title- Coreference Semantics from Web Features are-

1 . The Tradeoffs Between Open and Traditional Relation Extraction   P08-1004

2 . An Empirically Based System For Processing Definite Descriptions   J00-4003

3 . First-Order Probabilistic Models for Coreference Resolution   N07-1011

4 . Error-Driven Analysis of Challenges in Coreference Resolution   D13-1027

5 . Coreference Resolution across Corpora: Languages Coding Schemes and Preprocessing Information   P10-1144

2. KMeans

a. Papers Recommended for Paper ID- P10-1142

Title- Supervised Noun Phrase Coreference Research: The First Fifteen Years

1 . POS Tagging Versus Classes In Language Modeling   W98-1121

2 . Multilingual Aligned Parallel Treebank Corpus Reflecting Contextual Information And Its Applications   W04-2208

3 . Automatic Distinction Of Arguments And Modifiers: The Case Of Prepositional Phrases   W01-0715

4 . Discourse-New Detectors For Definite Description Resolution: A Survey And A Preliminary Proposal   W04-0707

5 . Extracting And Evaluating General World Knowledge From The Brown Corpus   W03-0902

b. Papers Recommended for Paper ID- P12-1041

Title- Coreference Semantics from Web Features

1 . Extracting Bacteria Biotopes with Semi-supervised Named Entity Recognition and Coreference Resolution   W11-1814

2 . A Clustering Approach For Unsupervised Chinese Coreference Resolution   W06-0106

3 . Extending English ACE 2005 Corpus Annotation with Ground-truth Links to Wikipedia   W10-3503

4 . The Impact Of Morphological Stemming On Arabic Mention Detection And Coreference Resolution   W05-0709

5 . A Chain-starting Classifier of Definite NPs in Spanish   E09-3006

3. <u>Kernel K Means</u>
   a. Papers Recommended for Paper ID- P10-1142

   Title- Supervised Noun Phrase Coreference Research: The First Fifteen Years

   1 . POS Tagging Versus Classes In Language Modeling   W98-1121

   2 . Multilingual Aligned Parallel Treebank Corpus Reflecting Contextual Information And Its Applications   W04-2208

   3 . Automatic Distinction Of Arguments And Modifiers: The Case Of Prepositional Phrases   W01-0715

   4 . The Uptake Of Online Tools And Web-Based Language Resources By Freelance Translators: Implications For Translator Training Professional Development And Research   W04-1406

   5 . The Treatment Of Noun Phrase Queries In A Natural Language Database Access System   W98-0606

   b. Papers Recommended for Paper ID- P12-1041

   Title- Coreference Semantics from Web Features

   1 . Extracting Bacteria Biotopes with Semi-supervised Named Entity Recognition and Coreference Resolution   W11-1814

   2 . A Clustering Approach For Unsupervised Chinese Coreference Resolution   W06-0106

   3 . A System For Extraction Of Temporal Expressions From French Texts Based On Syntactic And Semantic Constraints   W01-1314

   4 . Bootstrapping Parallel Treebanks   W04-1910

   5 . Development Of The Concept Dictionary Implementation Of Lexical Knowledge   W91-0219

4. <u>Subspace Clustering</u>
   a. Papers Recommended for Paper ID- P12-1041

   Title- Coreference Semantics from Web Features

   1 . Discriminative Reordering with Chinese Grammatical Relations Features   W09-2307

   2 . Non-Classical Lexical Semantic Relations   W04-2607

   3 . A System For Extraction Of Temporal Expressions From French Texts Based On Syntactic And Semantic Constraints   W01-1314

   4 . Bootstrapping Parallel Treebanks   W04-1910

   5 . Development Of The Concept Dictionary Implementation Of Lexical Knowledge   W91-0219

   b. Papers Recommended for Paper - P10-1142

   Title- Supervised Noun Phrase Coreference Research: The First Fifteen Years

   1 . Zero Anaphors In Chinese Discourse Processing   W00-1221

2 . Defining a Core Body of Knowledge for the Introductory Computational Linguistics Curriculum   W08-0204

3 . Thai WordNet Construction   W09-3420

4 . The Uptake Of Online Tools And Web-Based Language Resources By Freelance Translators: Implications For Translator Training Professional Development And Research   W04-1406

5 . The Treatment Of Noun Phrase Queries In A Natural Language Database Access System   W98-0606

5. Latent Factor Model

   a. **Papers Recommended for Paper - P12-1041**

   Coreference Semantics from Web Features :

   1 . Automatic Acquisition Of Hyponyms From Large Text Corpora   C92-2082

   2 . A Machine Learning Approach To Coreference Resolution Of Noun Phrases   J01-4004

   3 . Automatic Retrieval and Clustering of Similar Words   P98-2127

   4 . Improving Machine Learning Approaches To Coreference Resolution   P02-1014

   5 . A Model-Theoretic Coreference Scoring Scheme   M95-1005

   6 . Class-Based N-Gram Models Of Natural Language   J92-4003

   7 . Unsupervised Word Sense Disambiguation Rivaling Supervised Methods   P95-1026

   8 . Accurate Methods For The Statistics Of Surprise And Coincidence   J93-1003

   9 . Automatic Word Sense Discrimination   J98-1004

   10 . The Mathematics Of Statistical Machine Translation: Parameter Estimation   J93-2003

   b. **Papers Recommended for Paper - P10-1142**

   Supervised Noun Phrase Coreference Research: The First Fifteen Years :

   1 . Attention Intentions And The Structure Of Discourse   J86-3001

   2 . A Machine Learning Approach To Coreference Resolution Of Noun Phrases   J01-4004

   3 . Improving Machine Learning Approaches To Coreference Resolution   P02-1014

   4 . A Model-Theoretic Coreference Scoring Scheme   M95-1005

   5 . Centering: A Framework For Modeling The Local Coherence Of Discourse   J95-2003

   6 . Building A Large Annotated Corpus Of English: The Penn Treebank   J93-2004

   7 . An Algorithm For Pronominal Anaphora Resolution   J94-4002

8 . On Coreference Resolution Performance Metrics   H05-1004

9 . The Mathematics Of Statistical Machine Translation: Parameter Estimation J93-2003

MAE= 0.9144300070191597

6. [Nuclear Norm Minimization](#)

   a. **Papers Recommended for Paper - P10-1142**

   Title: Supervised Noun Phrase Coreference Research: The First Fifteen Years

   1 . Thumbs Up? Sentiment Classification Using Machine Learning Techniques W02-1011

   2 . Thumbs Up Or Thumbs Down? Semantic Orientation Applied To Unsupervised Classification Of Reviews   P02-1053

   3 . A Maximum-Entropy-Inspired Parser   A00-2018

   4 . Three Generative Lexicalized Models For Statistical Parsing   P97-1003

   5 . Recognizing Contextual Polarity In Phrase-Level Sentiment Analysis H05-1044

   MAE= 0.9410057256206139

   b. **Papers Recommended for Paper ID- P12-1041**

   Title: Coreference Semantics from Web Features :

   1 . Thumbs Up Or Thumbs Down? Semantic Orientation Applied To Unsupervised Classification Of Reviews   P02-1053

   2 . Discriminative Training Methods For Hidden Markov Models: Theory And Experiments With Perceptron Algorithms   W02-1001

   3 . Recognizing Contextual Polarity In Phrase-Level Sentiment Analysis H05-1044

   4 . Coarse-To-Fine N-Best Parsing And MaxEnt Discriminative Reranking P05-1022

   5 . Online Large-Margin Training Of Dependency Parsers   P05-1012

7. [Binary Matrix Factorisation](#)

   a. **Papers Recommended for Paper - P10-1142**

   Supervised Noun Phrase Coreference Research: The First Fifteen Years :

    are-

   1 . Building A Large Annotated Corpus Of English: The Penn Treebank J93-2004

   2 . The Mathematics Of Statistical Machine Translation: Parameter Estimation   J93-2003

   3 . Minimum Error Rate Training In Statistical Machine Translation P03-1021

4 . A Systematic Comparison Of Various Statistical Alignment Models
J03-1002

5 . Moses: Open Source Toolkit for Statistical Machine Translation
P07-2045

8. [CCIDF- Co-citation Inverse Document Frequency](#)

   a. Papers Recommended for Paper - P12-1041

Title: Coreference Semantics from Web Features :

1 . A Large-Scale Exploration Of Effective Global Features For A Joint Entity Detection And Tracking Model - H05-1013

2 . Coreference Resolution across Corpora: Languages Coding Schemes and Preprocessing Information - P10-1144

3 . Creating Robust Supervised Classifiers via Web-Scale N-Gram Data - P10-1089

4 . Understanding the Value of Features for Coreference Resolution - D08-1031

5 . Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art - P09-1074

   b. Papers Recommended for Paper - P10-1142

Title: Supervised Noun Phrase Coreference Research: The First Fifteen Years

1 . Improving Pronoun Resolution Using Statistics-Based Semantic Compatibility Information - P05-1021

2 . Capturing Salience with a Trainable Cache Model for Zero-anaphora Resolution - P09-1073

3 . Accurate Semantic Class Classifier for Coreference Resolution - D09-1128

4 . Competitive Self-Trained Pronoun Interpretation - N04-4009

5 . Corpus-Based Learning For Noun Phrase Coreference Resolution - W99-0634

Comparison Results
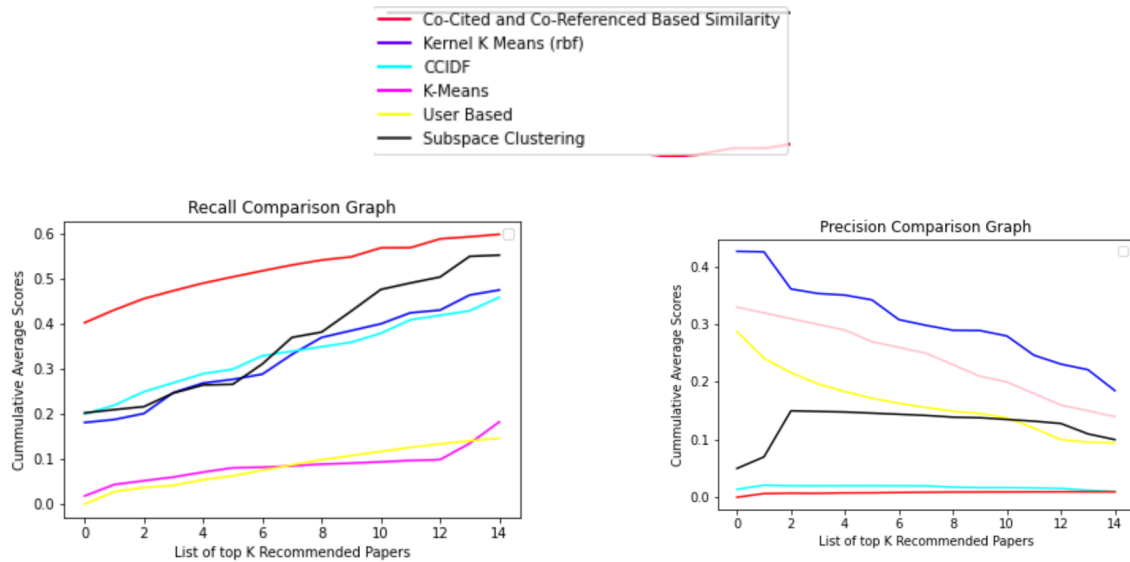
Matrix Factorization Based Approches
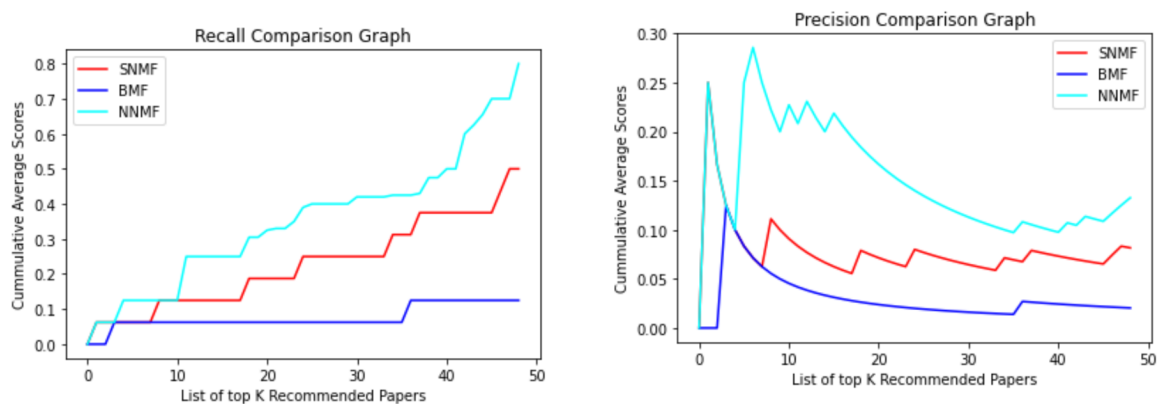
References
The following Papers were used to implement the project

1. Candidate Paper Based Methodology
   a. N. Sakib, R. B. Ahmad and K. Haruna, "A Collaborative Approach Toward Scientific Paper Recommendation Using Citation Context," in IEEE Access, vol. 8, pp. 51246-51255, 2020, doi: 10.1109/ACCESS.2020.2980589.
2. CC-IDF
   a. W. Tanner, E. Akbas and M. Hasan, "Paper Recommendation Based on Citation Relation," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 3053-3059, doi: 10.1109/BigData47090.2019.9006200.

    b. Beel, Joeran & Breitinger, Corinna & Langer, Stefan. (2017). Evaluating the CC-IDF citation-weighting scheme: How effectively can 'Inverse Document Frequency' (IDF) be applied to references?.

3. Citation Based Recommendation Evaluation Methodology
    a. H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele and F. Xia, "Context-Based Collaborative Filtering for Citation Recommendation," in IEEE Access, vol. 3, pp. 1695-1703, 2015, doi: 10.1109/ACCESS.2015.2481320.

4. Tosyali, A., Kim, J., Choi, J. *et al.* New node anomaly detection algorithm based on nonnegative matrix factorization for directed citation networks. *Ann Oper Res* 288, 457–474 (2020). https://doi.org/10.1007/s10479-019-03508-4